Capítulo 4

Google Hacking

4.1. Objetivos

- Entender o que é Google Hacking
- Conhecer os riscos que o Google traz
- Aprender como usar o Google como ferramenta auxiliar para um pentest
- Conhecer os principais comandos do Google
- Aprender como encontrar buscas pré-definidas, utilizando o GHD

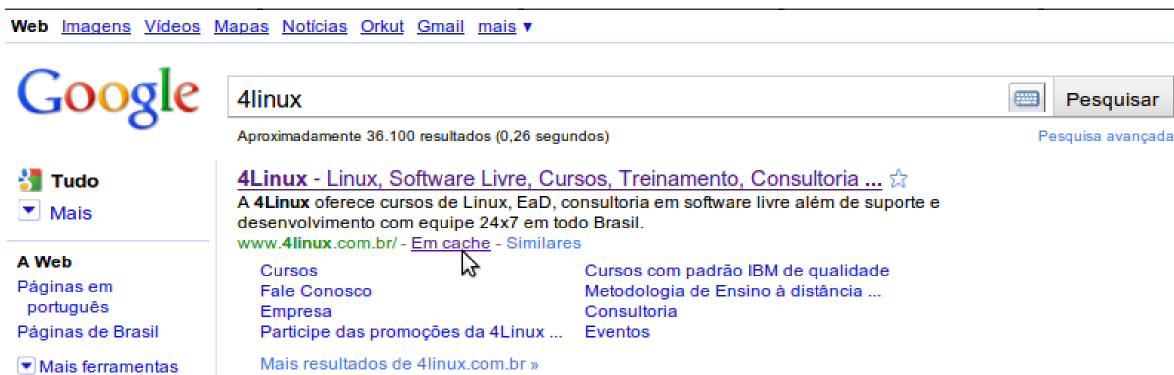
4.2. Google Hacking

Google Hacking é a atividade de usar recursos de busca do site, visando atacar ou proteger melhor as informações de uma empresa. As informações disponíveis nos servidores web da empresa provavelmente estarão nas bases de dados do Google.

Um servidor mal configurado pode expor diversas informações da empresa no Google. Não é difícil conseguir acesso a arquivos de base de dados de sites através do Google.

O Google possui diversos recursos que podem ser utilizados durante um teste de invasão, e justamente por isso é considerada a melhor ferramenta para os hackers, pois permite acesso a todo e qualquer tipo de informação que se queira.

Podemos usar como exemplo, o recurso de "cache" do Google, onde o mesmo armazena versões mais antigas de todos os sites que um dia j[á foram indexados por seus robôs.



Esse recurso permite que tenhamos acesso às páginas que já foram tiradas do ar, desde que ainda existam na base de dados do Google. Vamos imaginar que em algum momento da história do site de uma organização, uma informação mais sensível estivesse disponível. Depois de um tempo, o webmaster tendo sido alertado retirou tal informação do site. No entanto, se a página do site já tiver sido indexada pelo Google, é possível que mesmo tendo sido alterada, ou retirada, ainda possamos acessá-la utilizando o recurso de cache do Google.

4.3. Comandos Avançados do Google

4.3.1. intitle, allintitle

Busca conteúdo no título (tag title) da página.

Quando utilizamos o comando *intitle*, é importante prestar atenção à sintaxe da string de busca, posto que a palavra que segue logo após o comando *intitle* é considerada como a string de busca. O comando *allintitle* quebra essa regra, dizendo ao Google que todas as palavras que seguem devem ser encontradas no title da página, por isso, esse último comando é mais restritivo.

4.3.2. inurl, allinurl

Encontra texto em uma URL.

Como explicado no operador *intitle*, pode parecer uma tarefa relativamente simples utilizar o operador *inurl* sem dar maior atenção ao mesmo. Mas devemos ter em mente que uma URL é mais complicada do que um simples title, e o funcionamento do operador *inurl* pode ser igualmente complexo.

Assim como o operador *intitle*, *inurl* também possui um operador companheiro, que é o *allinurl*, que funciona de maneira idêntica e de forma restritiva, exibindo resultados apenas em que todas as strings foram encontradas.

4.3.3. filetype

Busca por um arquivo de determinado tipo.

O Google pesquisa mais do que apenas páginas web. É possível pesquisar muitos tipos diferentes de arquivos, incluindo PDF (Adobe Portable Document Format) e Microsoft Office. O operador filetype pode ajudá-lo na busca de tipo de arquivos específicos. Mais especificamente, podemos utilizar esse operador para pesquisas de páginas que terminam em uma determinada extensão.

4.3.4. allintext

Localiza uma string dentro do texto de uma página.

O operador *allintext* é talvez o mais simples de usar, pois realiza a função de busca mais conhecida como: localize o termo no texto da página.

Embora este operador possa parecer genérico para ser utilizado, é de grande ajuda quando sabe que a string de busca apenas poderá ser encontrada no texto da página. Utilizar o operador *allintext* também pode servir como um atalho para "encontrar esta string em qualquer lugar, exceto no title, URL e links".

4.3.5. site

Direciona a pesquisa para o conteúdo de um determinado site.

Apesar de ser tecnicamente uma parte da URL, o endereço (ou nome de domínio) de um servidor pode ser mais bem pesquisada com o operador *site*. *Site* permite que você procure apenas as páginas que estão hospedadas em um servidor ou domínio específico.

4.3.6. link

Busca por links para uma determinada página.

Em vez de fornecer um termo de pesquisa, o operador necessita de um link URL ou nome do servidor como um argumento.

4.3.7. inanchor

Localiza texto dentro de uma âncora de texto.

Este operador pode ser considerado um companheiro para o operador link, uma vez que ambos buscam links. O operado inanchor, no entanto, pesquisa a representação de texto de um link, não o URL atual.

Inanchor aceita uma palavra ou expressão como argumento, como

inanchor:click ou inanchor:oys. Este tipo de pesquisa será útil especialmente quando começamos a estudar formas de buscar relações entre sites.

4.3.8. daterange

Busca por páginas publicadas dentro de um “range” de datas.

Você pode usar este operador para localizar páginas indexadas pelo Google em um determinado intervalo de datas. Toda vez que o Google rastreia uma página, a data em sua base de dados é alterada. Se o Google localizar alguma página Web obscura, pode acontecer de indexá-la apenas uma vez e nunca retornar à ela.

Se você achar que suas pesquisas estão entupidas com esses tipos de páginas obscuras, você pode removê-las de sua pesquisa (e obter resultados mais atualizados) através do uso eficaz do operador daterange.

Lembrando que a data deve ser informada no formato do calendário Juliano, informando o número de dias existentes entre 4713 AC e a data em que se quer buscar.

4.3.9. cache

Mostra a versão em cache de uma determinada página.

Como já discutimos, o Google mantém "snapshots" de páginas que indexou e que podemos acessar através do link em cache na página de resultados de busca. Se quiser ir direto para a versão em cache de uma página, sem antes fazer uma consulta ao Google para chegar ao link em cache na página de resultados, você pode simplesmente usar o operador *cache* em uma consulta, como *cache*:blackhat.com ou *cache*:*www.netsec.net/content/index.jsp*.

4.3.10. info

Mostra conteúdo existente no sumário de informações do Google.

O operador info mostra o resumo das informações de um site e fornece links para outras pesquisas do Google que podem pertencer a este site. O parâmetro informado à este operador, deve ser uma URL válida.

4.3.11. related

Mostra sites relacionados.

O operador related exibe o que o Google determinou como relacionado a um determinado site. O parâmetro para esse operador é uma URL válida. É possível conseguir essa mesma funcionalidade, clicando no link "Similar Pages" a partir de qualquer página de resultados de busca, ou usando o "Find pages similar to the page" da página do formulário de pesquisa avançada



Dica: Se você está realizando um pentest em um site chinês, para que usar um google.com.br? O Google prioriza os resultados para determinados websites. Raramente você vê páginas escritas em japonês, chinês, árabe e outros quando usa o google.com.br, não? Uma boa busca é feita em servidores diferentes, com países diferentes.

4.4. Google Hacking Database

Há um banco de dados virtual, com tags de busca no Google previamente criadas, para conseguir informações específicas.

A partir das tags existentes, podemos encontrar muitas coisas interessantes sem precisarmos nos preocupar em como desenvolver buscas específicas, utilizando os operadores do Google, e testá-las até conseguirmos que os filtros corretos funcionem.

Mas o mais importante que devemos manter em mente, é a possibilidade e adaptar tais tags de busca para nossas necessidades.



Google Hacking Database: http://johnny.ihackstuff.com/ghdb/

4.5. Levantamento de informações

O Google é a principal ferramenta para o levantamento de informações de nosso alvo. É o melhor sistema público para utilizarmos em busca de informações sobre qualquer coisa em relação ao nosso alvo: sites, propagandas, parceiros, redes sociais, grupos e etc.

Além do Google, há outros sites específicos que auxiliam no processo de levantamento de informações, os quais conheceremos mais adiante.

Um simples exemplo do que podemos encontrar no Google, e que pode voltar-se contra a pessoa que disponibilizou tais informações online, é o seguinte: digitar na caixa de busca currículo + cpf .

Certamente vários resultados retornarão com links onde podemos encontrar nome completo, endereço, telefone, CPF, identidade e mais algumas informações das pessoas que disponibilizaram seus dados na internet. Tendo conhecimento de como esses dados podem ser utilizados de maneira maliciosa, podemos ter mais consciência ao publicarmos quaisquer informações nossas na internet.

4.6. Contramedidas

- Possuir uma boa política referente às publicações de informações na internet.
- Não deixar configurações padrão em servidores web, para que os mesmos não consigam ser identificados facilmente.
- Sempre analisar as informações disponíveis sobre a empresa em sites de busca.
- Alertar e treinar os funcionários da empresa com relação a maneira com que um ataque de engenharia social pode acontecer, e as possíveis informações que o atacante poderá usar nesse ataque.

4.7. Prática dirigida

Busca por arquivos de base de dados em sites do governo:

- site:gov.br ext:sql

Busca por um servidor específico

- inurl:"powered by" site:sistema.com.br

A pesquisa busca arquivos de e-mail em formato .mdb

- inurl:e-mail filetype:mdb

Essa pesquisa busca telefones disponíveis em intranet encontradas pelo Google

- inurl:intranet + intext:"telefone"

Realizando uma pesquisa dessa maneira é possível identificar muitos dos subdomínios da Oracle

- site:oracle.com -site:www.oracle.com

Detectando sistemas que usando a porta 8080

- inurl:8080 -intext:8080

Encontrando VNC

- intitle:VNC inurl:5800 intitle:VNC

Encontrando VNC

- intitle:"VNC Viewer for Java"

Encontrando Webcam ativa

- "Active Webcam Page" inurl:8080

Encontrando Webcam da toshiba:

- intitle:"toshiba network camera - User Login"

Encontrando Apache 1.3.20:

- "Apache/1.3.20 server at" intitle:index.of

Asterisk VOIP Flash Interface

- intitle:"Flash Operator Panel" -ext:php -wiki -cms -inurl:as

Possíveis falhas em aplicações web:

- allinurl:".php?site="
- allinurl:".php?do="
- allinurl:".php?content="
- allinurl:".php?meio="
- allinurl:".php?produto="
- allinurl:".php?cat="